

## Who Knows What, When? Current and Desired Capacities for Online Journal Statistics Gathering and Dissemination

James MacGregor  
*Simon Fraser University*  
Andrea Kosavic  
*York University*

Scholarly and Research  
Communication

VOLUME 1 / ISSUE 1 / 2010

### Abstract

As part of the national Synergies project, the Statistics Working Group was formed to investigate statistical reporting mechanisms used by participating institutions, to research online journal-specific reporting needs, and to form a common model for statistical reporting and the sharing of usage data across Canada. The working group informally compared the statistics-gathering range of Open Journal Systems (OJS) and the Érudit Consortium publishing platform; they also surveyed Canadian and international scholarly journal stakeholders to obtain a better understanding of their needs. Respondents were asked about desired types of statistics captured, preferred groupings, preferred harvesting frequency, and their level of satisfaction with available tools. This article describes the results of the platform comparison and the survey, and it provides a set of recommendations intended for the Synergies project but applicable elsewhere.

### Keywords

Synergies Canada; Statistics; Open Journal Systems; Open Conference Systems; Érudit; Open Access; Interoperability

**James MacGregor** is a member of the Synergies National Technical Committee and is a Public Knowledge Project Associate at Simon Fraser University Library, 8888 University Drive, Burnaby, BC V5A 1S6. Email: [jmacgreg@gmail.com](mailto:jmacgreg@gmail.com).

**Andrea Kosavic** is a member of the Synergies National Technical Committee and is the Digital Initiatives Librarian at York University, 4700 Keele Street, Toronto, ON M3J 1P3. Email: [akosavic@yorku.ca](mailto:akosavic@yorku.ca).

CCSP Press

*Scholarly and Research Communication*

Volume 1, Issue 1, Article ID 010201, 17 pages

Journal URL: [www.src-online.ca](http://www.src-online.ca)

Received October 13, 2009, Accepted November 25, 2009, Published January 7, 2010

MacGregor, James, & Kosavic, Andrea. (2010). Who Knows What, When? Current and Desired Capacities for Online Journal Statistics Gathering and Dissemination. *Scholarly and Research Communication*, 1(1): 010201, 17 pp.

© 2010 James MacGregor, Andrea Kosavic. This Open Access article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc-nd/2.5/ca>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## **Introduction**

Synergies is a scholarly Canadian collaboration whose primary aim is to support the publication and dissemination of social sciences and humanities research across Canada and internationally (Synergies, 2008). The organization is a consortium of academic institutions with major nodes based in the Atlantic provinces, Québec, Ontario, the Prairies, and British Columbia. Technical concerns related to the building of a national platform for research dissemination are relegated to a Technical Committee, whose members are drawn from all nodes. Working groups are composed of members of the Technical Committee and are assigned specific areas that require research and recommendations. These working groups are thus largely responsible for the visualization and development objectives of the Synergies platform. Since regional Synergies members include the developers of the Érudit publishing platform and Open Journal Systems (OJS), any technical recommendations put forward by these working groups mandate sometimes-significant development or technical changes in those software platforms.

The Statistics Working Group is composed of three members of the Technical Committee, each from different regional nodes. The group's goals are threefold: to determine statistical and reporting needs at the regional and national level as defined by the Synergies mandate; to determine the statistical and reporting needs of Synergies "stakeholders" (journals and their managers, repositories and their managers, libraries, regional platforms, et cetera); and to recommend how, and at which point, statistics will be gathered and displayed.

To best meet these goals, the group decided to complete two tasks concurrently: a) to examine statistical gathering and reporting methods currently available in the most common applications (Érudit and OJS); and b) to survey identified stakeholders to assess how they are currently capturing statistics, where their systems are not sufficient, and what kind of statistics they most need. We could then make recommendations based on the current picture (platform analysis), cross-referenced against what stakeholders report that they need.

## **Platform comparison**

The two primary platforms used by Synergies nodes are the Érudit publishing platform and OJS. The Érudit publishing platform is a single website and corresponding set of tools developed and maintained by the Érudit Consortium (<http://erudit.org>). It provides access to a multitude of scholarly material, including journal articles, monographs, theses, and proceedings. OJS is an open source journal management and publishing system developed by the Public Knowledge Project (PKP; <http://pkp.sfu.ca>) and used around the world.

Two other platforms – Open Conference Systems (OCS), also developed by the PKP, and DSpace – were initially included in the platform analysis but were deemed beyond the current scope of the working group. A fair portion of any analysis and the resulting recommendations toward OJS development may also be applicable to OCS; while an analysis of DSpace might provide interesting results, Synergies' capacity to commit development resources toward an external project is limited.

## METHODOLOGY

The working group identified a number of usage items and statistical types considered most likely relevant to stakeholders, and we analyzed the most recent incarnation of each platform. In the case of OJS, native reporting capabilities were recorded; special care was taken to also include and visibly note where third-party reporting mechanisms (Google Analytics and phpMyVisites) were available via plug-ins. Although OJS includes COUNTER support as a plug-in, it is not strictly a third-party tool and gathers its data directly from the OJS database; any features provided by it were checked as if part of the native system. The final analysis can be seen in the platform comparison chart (see Table 1).

In the case of Érudit, the analysis was not in fact completed against the Érudit platform itself, but against AWStats 6.7: Érudit does not include any native reporting tools and relies on AWStats for usage data. AWStats is a robust server log analysis tool that provides a plethora of usage data, from unique visitors and number of visits to page views; text of search queries; geographic, browser, and operating system information. It can also be used alongside PKP software, and in fact contains a vast amount of statistical information that OJS and OCS do not currently capture. Although some information AWStats tracks is doubtless of use to stakeholders, the manner in which it is currently displayed is far too detailed for most uses, and in some cases relevant information (for example, total article views) may be difficult to interpret or extract from the default report. This can also be said of other third-party reporting and/or analysis tools, such as Google Analytics, phpMyVisites, and Piwik.

**Table 1: Comparison chart of statistics natively gathered by OJS, OCS, Érudit and DSpace**

Statistic	Open Journal Systems	Open Conference Systems	Érudit	DSpace
<b>Abstract views</b>	captured in database, not captured in reports, no timestamp recorded - running total only	captured in database, not captured in reports, no timestamp recorded - running total only	queries in AWStats Pages-URL section	captured in reports, monthly totals, not object specific (called Item Views)
<b>PDF/html views</b>	captured in database, captured in COUNTER report, reported monthly	captured in database, reported in interface	queries in AWStats Pages-URL section	captured in reports, object specific, monthly totals (called Bitstream views)
<b>Issue views</b>	no*	n/a; the equivalent would be scheduled conference views, in which case the answer is no*	queries in AWStats Pages-URL section	the equivalent would be collection/ community views that are captured monthly
<b>Items added</b>	yes, by any specified period of time	yes, by any specified period of time	no	yes, by month

Page views	no*	no*	AWStats, for day/month	no
Hits	yes* via plug-ins: Google Analytics, phpMyVisites	no*	AWStats, for day/month	no
Bandwidth	no*	no*	AWStats, for day/month	no
Accounts created	yes, by any specified period of time - subscriptions and readers also counted	yes, by any specified period of time - registrations and readers also counted	no	yes, by month
User logins	no	no	no	yes, monthly total and broken down by individual user
Unique Visitors	yes* via plug-ins: Google Analytics, phpMyVisites	no*	AWStats, for day/month	no
Visits	yes* via plug-ins: Google Analytics, phpMyVisites	no*	AWStats, for day/month	no
Visit duration	yes* via plug-ins: Google Analytics, phpMyVisites	no*	AWStats by month	no
Institutional visits	yes* via plug-ins: Google Analytics, phpMyVisites	no*	Geo IP Maxmind plug-in for AWStats	no
Visitor domains/ countries	yes* via plug-ins: Google Analytics, phpMyVisites	no*	AWStats by month	no
Robots/ Spiders visits	yes* via plug-ins: Google Analytics, phpMyVisites	no*	AWStats by month	no
Hosts	yes* via plug-ins: Google Analytics, phpMyVisites	no*	AWStats by month	no
Search Keyphrases	these are visible in the log* but are difficult to count, cannot be organized or sorted. Can also be provided by plug-ins	visible in log; not displayed (organized or sorted)	AWStats by month	no
Search Keywords	these are visible in the log* but are difficult to count, cannot be organized or sorted. Can also be provided by plug-ins	visible in log; not displayed (organized or sorted)	AWStats	words searched by month (more than 3 times)

<b>Queries (required for ARL stats)</b>	these are visible in the log* but are difficult to count, cannot be organized or sorted	these are visible in the log* but are difficult to count, cannot be organized or sorted	no	total searches by month; also OAI requests by month
<b>Links from...</b>	no*	no*	AWStats by month	no
<b>File Types accessed</b>	yes* via COUNTER plug-in; only accounts for HTML/PDF galleys, does not account for other files	no*	AWStats by month	no
<b>Pages by URL</b>	yes* via plug-ins: Google Analytics, phpMyVisites	no*	AWStats by month	no
<b>Operating Systems</b>	yes* via plug-ins: Google Analytics, phpMyVisites	no*	AWStats by month	no
<b>Browsers</b>	yes* via plug-ins: Google Analytics, phpMyVisites	no*	AWStats by month	no
<b>Origin (linked from)</b>	yes* via plug-ins: Google Analytics, phpMyVisites	no*	AWStats by month	no
<b>Add to favourites</b>	yes* via plug-ins: Google Analytics, phpMyVisites	no*	AWStats by month	no
<b>HTTP Status Codes</b>	yes* via plug-ins: Google Analytics, phpMyVisites	no*	AWStats by month	no

\*OJS and OCS can also be reported on using AWstats, and can be assumed to support the same level of AWstats reporting as detailed in Érudit.

## PRELIMINARY OBSERVATIONS

It was noted that all participating Synergies institutions use either AWStats or Webalizer on their Web servers to glean usage statistics. Typically, external reporting programs analyze server log files for relevant information to display and do not provide the option of customized usage reports. In particular, Érudit relies exclusively on AWStats for usage reports, since the Érudit platform does not include any internal reporting functionality.

At the time of comparison, OJS included multiple points of statistical and usage data gathering:

- a basic and non-compliant COUNTER Release 2 report plug-in (since upgraded to Release 3 and made more compliant), used to report Full-Text PDF and/or HTML article view counts on a per-month basis across an entire installation, in csv format only
- a Google Analytics plug-in as well as a phpMyVisites (now known as Piwik) plug-in, both of which could be used for usage reporting, in a manner similar to AWStats or Webalizer

- two report plug-ins that export information in comma-separated value (csv) format: an Articles report, which generates a list of articles and relevant information on a per-journal basis; and a Review report, which generates a list of review assignments in the journal
- access to journal-level statistics (number of issues published, number of items published, number of total submissions, et cetera) for Journal Managers on a per-year basis, which can optionally be made publicly viewable
- individual view counts for article galleys (e.g., PDF, HTML), available from an individual submission's Editing page

OJS' plug-in structure allows users to extend the system without making changes to core files. The Article and Review report plug-ins are examples of the plug-in framework's reporting capabilities, which can be extended for other types of reports.

### PRELIMINARY ANALYSIS

A preliminary analysis of the comparison chart, supported by general familiarity with each system, shows that neither platform shares a common statistical collection mechanism. While OJS can natively capture the number of item record/abstract views and the text of search queries, other statistics are currently difficult to capture without using third-party tools. This third-party tool problem is especially pertinent to Érudit, which relies exclusively on AWStats. As has been stated above: although these tools may provide relevant statistics, these numbers are buried among a large, superfluous amount of other information and may be difficult to interpret or extract; furthermore, although these tools may in some cases provide custom report functionality, developing custom third-party reports would be time-consuming and would likely result in ongoing maintenance issues. There are also concerns regarding third-party tools that store and process usage data on their own servers, for example Google Analytics: while usage data may not necessarily be confidential, ceding storage and control responsibilities to commercial and to some degree unaccountable companies may be unacceptable. In some cases, this issue is exacerbated by the companies in question storing data outside of Canada.

During a committee meeting, the Technical Committee agreed that while the information gathered by these tools may be useful to individual journals or organizations, from the Synergies standpoint these tools gathered far more information than necessary, and that the necessary information should be gathered in a way that was easier to control, maintain, and share, that is, within the systems themselves.

Further analysis pointed out that there was no easily discernible standard type of report (i.e., XML or CSV) or mechanism for sharing reports across the platforms. To share relevant reports back and forth – and, most significantly, to send reports from the regional nodes to the head node and back – a common report type, and a mechanism to share said reports, would be of inestimable usefulness.

A congruent but separate investigation of the COUNTER protocol and its use in OJS prompted a look at the Standardized Usage Statistics Harvesting Initiative (SUSHI) protocol for the purpose of moving usage reports from the regional level to the head node on

an as-needed basis. The COUNTER (Counting Online Usage of Networked Electronic Resources) Project is an international initiative whose aim is to develop credible usage and statistics reports for different types of electronic resources, such as journal articles and database views. The project periodically releases Codes of Practice, which include detailed information on what kinds of usage data are to be gathered and disseminated, and how. This is presented as a set of protocols for different data types: journals, databases, books, and reference works. The most recent Code of Practice is Release 3, and it includes rules for formatting reports in both XML and CSV format (COUNTER, 2008). The SUSHI standard is an XML SOAP (Simple Object Access Protocol) specification to facilitate the transfer of COUNTER and other XML-based usage reports from one platform to another, typically an electronic resource management (ERM) system (SUSHI, 2008).

Some elements of the SUSHI protocol stand out as particularly useful in a Synergies context:

- SUSHI does not support authentication explicitly, but it can be implemented with authentication support (SUSHI 2008).
- The technology behind it (SOAP transactions and XML) is well understood by the Synergies Technical Committee and project developers, and would be easy to implement and maintain.
- The resulting code/protocol can also be used in conjunction with other services, ERMS, et cetera, at local and aggregator levels, as need be.
- SUSHI can be used to transmit all manner of usage reports, not just COUNTER, so long as the report is available in XML (SUSHI, 2008).

That last point in particular is important: the COUNTER reports only provide basic usage metrics, such as article full-text view counts (JR1) and database queries (DBR1) (COUNTER, 2008), and by no means do they provide all metrics presumed necessary to stakeholders.

## **PRELIMINARY RECOMMENDATION**

The preliminary recommendations from this stage of our research are to use SUSHI to share statistics between nodes; to implement, as a test case and baseline reporting mechanism, the COUNTER JR 1 report up to the Release 3 standard, which would then include both XML and CSV output; and to aim to extend the existing COUNTER report to satisfy our stakeholders.

## **Survey**

### **METHODOLOGY**

Early in our preliminary research, it became clear that it would be useful to better understand the statistical capture requirements of journal stakeholders as a whole, to best anticipate the needs of our current and future user community. Via direct email, we distributed a survey geared toward journal managers and editors to the international journal contact list maintained by the Public Knowledge Project, journals affiliated with Synergies partner institutions, and journals registered with the Directory of Open Access Journals, where contact information was made available on their respective websites.

Contextual information was requested from each respondent, including where and on what platform the journal is hosted, their subject areas of focus, and whether the journal



is open access. In addition, the survey gathered information about the methodology of statistics capture, the sharing of statistics, and the formatting of statistical reports. A detailed set of questions focusing on the capture of specific statistics, statistical groupings, and frequency of capture was also included. It was hoped that the survey would help the working group determine what were the most important statistics to capture and how users would like them to be shared and displayed.

The survey was distributed internationally; however, we did promote the survey more heavily to local Synergies partners and Canadian contacts, to ensure a minimum number of Canadian respondents to represent the needs of our future user community.

#### **OBSERVATIONS AND ANALYSIS**

The survey received responses from 243 respondents; 67.5% of these respondents completed the entire survey. The survey was made available in English and French, and 15% of our respondents completed the French version of the survey. Replies representing journals based in Canada formed 31% of our data.

We were pleased to see a good distribution of representation from various subject areas. Respondents were able to indicate percentages of subject area coverage for their journals, and when these were totalled, the subject area distribution was 35% science, 25% humanities, 34% social sciences, and 6% other.

Although 80% of respondents reported the adoption of an open access policy for their journals, there was a marked difference between Canadian and international respondents. For example, 48% of Canadian respondents represented open access journals, compared with 94% of international respondents. The remaining Canadian content was split between restricted access models at 11% and embargoed content at 28%.

The difference between international and Canadian journal open access adoption can likely be explained by the attempts of Synergies partner institutions to directly solicit journals from as many Canadian journals as possible, while international respondents were solicited largely from journals listed on the Directory of Open Access Journals, contacts from the Public Knowledge Project, and from international mailing lists. As a result, our Canadian data is likely a more comprehensive sampling of national open access adoption of online journals.

Academic institutions were the largest hosting body for journal content, dominating the category with 58% of the distribution. Scholarly societies formed the second largest group at 12%, followed by 10% with commercial hosting services and 8% with academic libraries.

OJS was the most popular hosting platform, with 35% of respondents reporting the use of this system. Canadian adoption of OJS was higher, at 44%, versus international adoption, at 30%. (It should again be noted that a number of journal contacts were culled from PKP-specific resources, so OJS would naturally be favoured.) In second place were locally designed systems, at 26% overall. It should be noted that over 30% of overall respondents were unable to specify software details or cited responses that fell outside of the categories provided as possible answers.



The high percentage of locally designed systems, at 26%, was somewhat surprising, as there are multiple open source systems available free of cost that are designed to facilitate the hosting of electronic journals. It would be interesting to investigate whether a correlation exists between the date of the journal being made available digitally and whether the system is locally designed. We did not capture the dates that journals were made available online and hence could not determine if most locally designed systems were launched prior to the rise in popularity of systems such as OJS.

The category of “locally designed system” as it appears in the survey is problematic in retrospect, as we cannot specify how users interpreted this category. For example, we are unable to answer precisely how many of these locally designed systems were created from scratch versus how many had adopted existing software such as Drupal or Wordpress to display their journals online. Ideally, subcategories of locally designed systems would have been made available in the survey.

The remainder of the survey raised questions concerning statistics collection practices. As OJS users ourselves, we are conscious of our use of more than one statistics-gathering method, and we were interested to see whether this was a widespread phenomenon. We also analyzed the responses to determine whether certain practices or software solutions emerged as more widely adopted.

A variety of statistics-gathering solutions was observed. Overall, the most popular tools used were built-in tools provided by the software platform (29%) and server log analyzers such as Webalizer and AWStats (26%). Twenty-two percent of respondents were unsure of statistics collection methods being used, and about 20% of users were also making use of free page-tagging services such as Google Analytics. It was interesting to note that 14% of respondents were also using informal, manual reporting solutions.

Seeing this wide variety of statistics collection methods, we mined our data to determine what percentage of our respondents were making use of more than one solution to capture statistical information about their journal. We discovered that approximately 30% of our overall respondents were making use of two or more statistics collection methods.

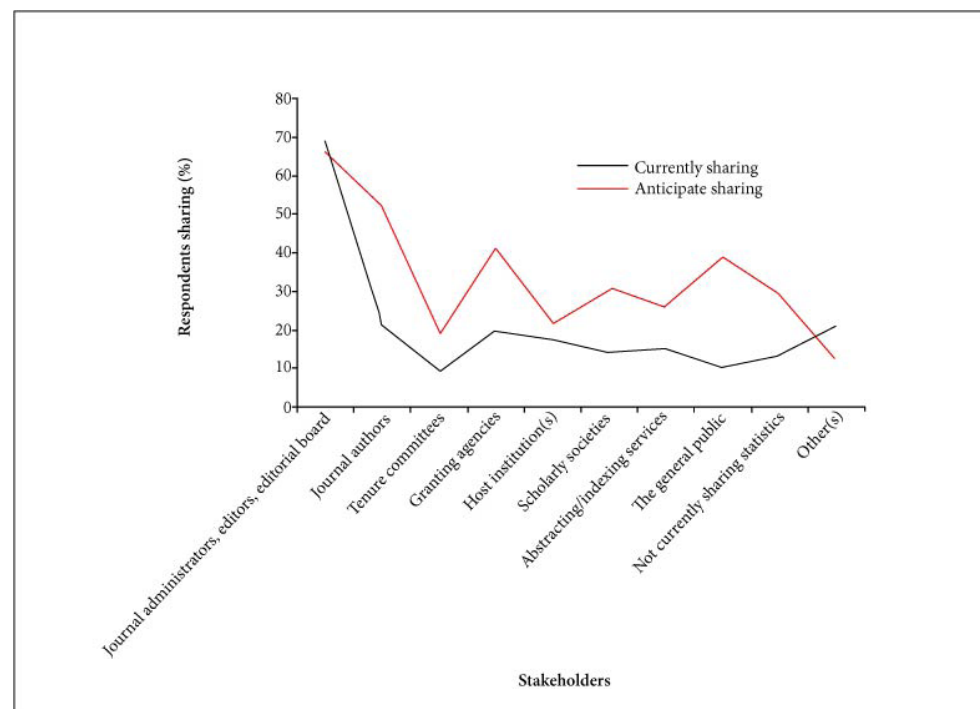
When filtering for all the OJS users that responded to this question, which were over a quarter of all respondents, we found that 52% of OJS users reported using the built-in statistics collection tools. This subgroup of respondents were also using a variety of other statistics collection methods, such as free page tagging services, manual reporting, and server log analyzers. For this OJS user group, we found that 38% of these users were also using two or more statistics collection methods.

We expected the above two numbers to be higher than 30% (general) and 38% (OJS-specific), as we have informally observed that our Canadian colleagues employ multiple statistics collection solutions. Nevertheless, both numbers demonstrate a need for flexibility when considering statistics solutions for the Synergies portal.

Stakeholders were asked if they were satisfied with the statistics reporting tools they are currently using. Due to the free-text nature of the response to this question, out of 59 respondents, only 35 answers could be interpreted as a clear yes or no. Most of the remaining answers elaborated on their statistical setup without indicating whether they were satisfied with it. It was interesting to observe the variety of tools being used, which included locally developed solutions; paid services such as Google's Urchin, Webtrends, and Extreme Tracking; and free tools such as Site Meter. For those who did indicate their level of satisfaction, 26% indicated that they were satisfied with their current tools, while the remaining 74% were not.

Sharing of statistics was also a topic of investigation. We asked our respondents to list the stakeholders with which they currently share their statistics and with which they anticipate sharing their statistics (Figure 1).

**Figure 1: Current and anticipated statistics sharing practices with stakeholders**



As anticipated, journal management stakeholders were the largest group, at 70%, for current sharing practices, and this percentage remained constant for anticipated sharing practices. There was, however, a noticeable gap between current and anticipated sharing practices for some stakeholder groups, where current sharing procedures were lower than anticipated sharing practices. This gap was largest for granting agencies and the general public.

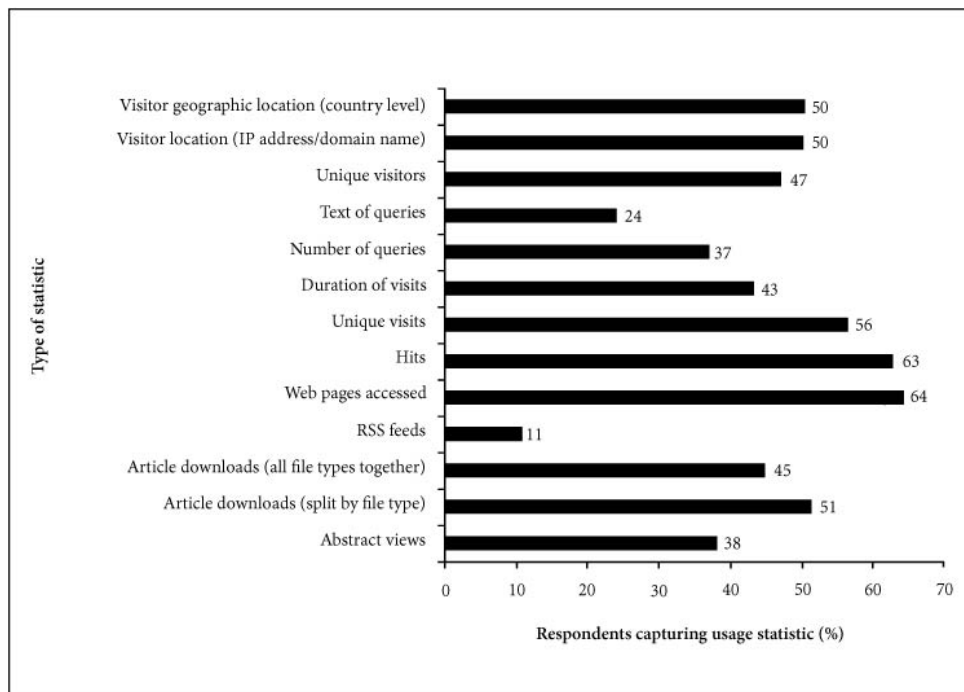
We do not know whether sharing statistical information with the public is restricted due to system limitations, as this was not investigated in the survey. It would be useful to observe whether the availability of a statistics tool that enables site managers to fine-tune which statistics are visible to the public would close the gap between current

and anticipated statistics sharing with the general public. As members of the Statistics Working Group we are challenged to be mindful of this gap and work toward a platform that minimizes the obstacles to data sharing.

It is of interest to discover whether the gap between current versus anticipated sharing of statistics with abstracting and indexing agencies may be due to the age of the journal. The H. W. Wilson Company, for example, when considering a title for abstracting, looks at quality indicators such as impact factor, wide holdings among relevant academic libraries, and recommendation by standard reference sources (Mark & Tarullo, 2003). For a journal to meet these criteria, it requires time to establish itself. We were unable to determine whether the gap between current and anticipated sharing of statistics with abstracting and indexing agencies was higher for younger journals, as we did not capture the date of journal establishment from our respondents.

Data was collected about which statistics were currently being captured by our respondents (Figure 2). Many of the respondents were unsure of the statistics captured by their journal websites; these replies were excluded from Figure 2. The most commonly collected statistics were basic data about site access, such as the number of times Web pages were accessed (64% of all respondents) and number of site hits (63%). Data about visitors was also frequently collected, which includes number of unique visits to the site (56%) and the geographic and Internet location of visitors (50%). The least collected statistics were the number of subscribers to the site's RSS feed (11%) and the text of queries inputted by users when searching the website (24%).

**Figure 2: Usage statistics captured by respondents**

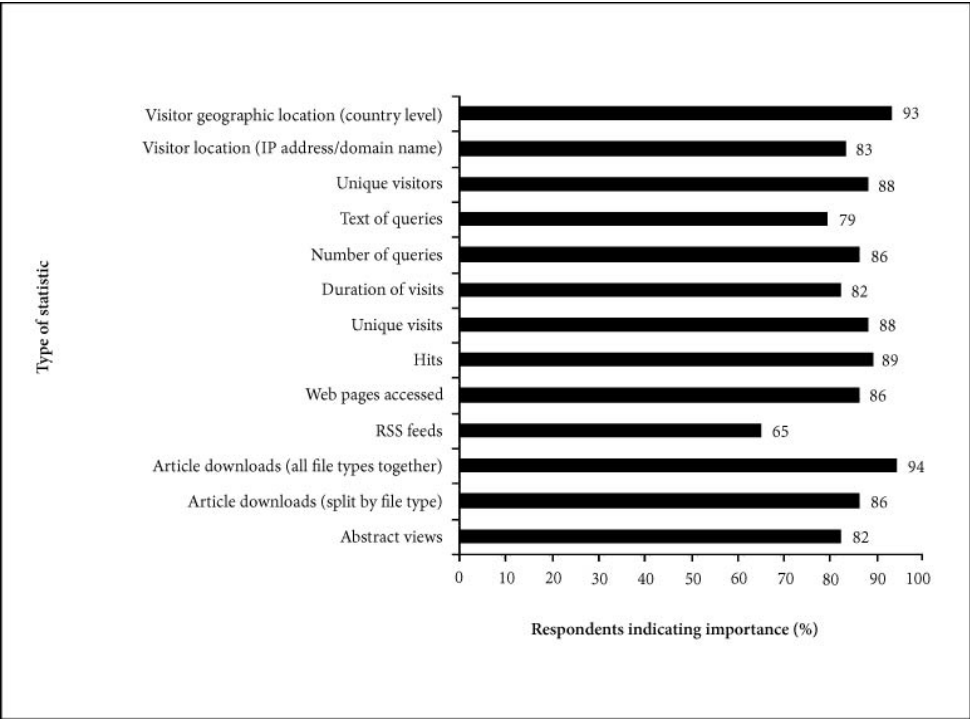


The number of respondents collecting statistics on article downloads by file type (51% of all respondents) and abstract views (38%) was lower than expected. To check

for accuracy of respondent replies to this question, we filtered out all OJS users and examined their responses to this question, as we know that OJS collects article download information via COUNTER plug-in. We were surprised to see that only 54% of OJS respondents replied that their systems collected statistics on article downloads. The fact that 46% our respondents using OJS were not fully aware of the statistics capabilities of their systems suggests that the low article download statistics cannot be taken at face value.

Respondents were also asked to indicate which statistics they view as important to capture (Figure 3). The statistics that emerged as most important to capture were the geographic location of visitors and the number of article downloads. Respondents were also interested in number of unique visits and visitors, and number of hits. Although RSS feed counts were perceived as the least important statistic for capture, there was a large gap between those currently capturing RSS feeds (11%) and those that viewed RSS feed counts as important to capture (65%).

Figure 3: Usage statistics perceived as important to capture



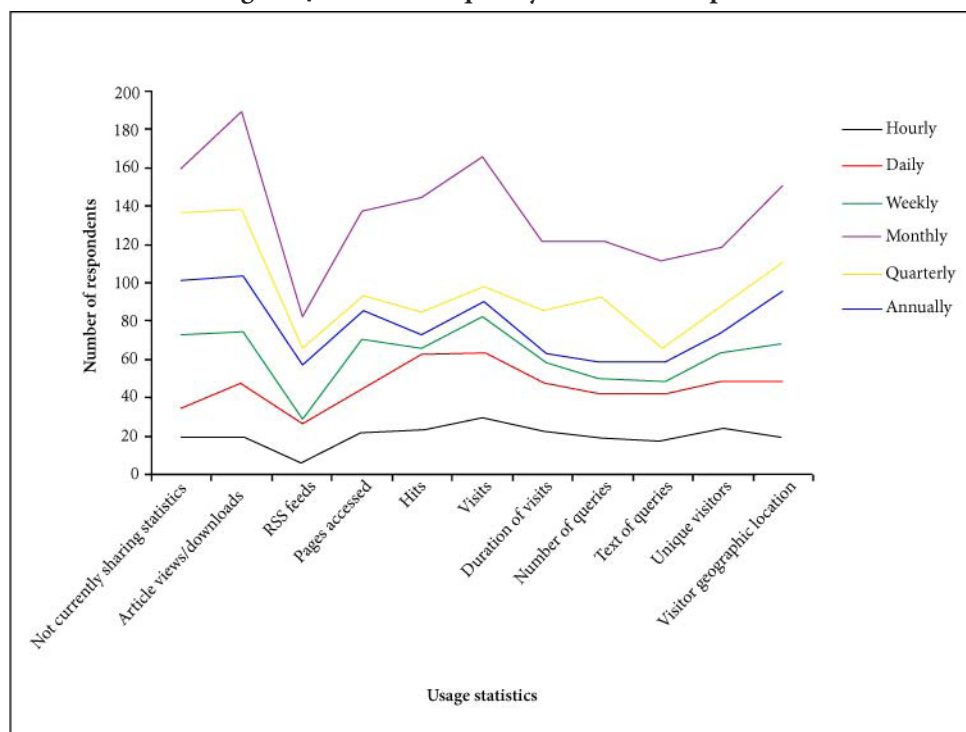
Although we have shown that a large percentage of OJS users are not fully knowledgeable about the statistics-collecting capabilities of their systems, we must not discount the possibility that some of the respondents may be lacking the technical infrastructure to collect information on article downloads. Since respondents indicated that the number of article downloads was the single most important statistic to capture, from a Synergies perspective we must ensure that article download information is captured by all of our partners.

It was important for us to understand the frequency with which journal stakeholders are interested in viewing statistical information. This will help to give us some baseline goals to meet for statistics capture and exchange at the Synergies portal level. We see from Figure 4 that a monthly capture of statistics is most essential.

When asked if it would be useful to view statistics generally grouped by article, author, journal section, or date span, respondents showed the highest preference for groupings by article and also showed an interest in groupings by author. There was very little interest in groupings by journal section or date span.

We also asked respondents to indicate interest in more detailed types of groupings, such as lists of top ten authors and top viewing countries. We were inspired to ask this question from observing these statistical groupings offered by repository platforms. PKP has also received requests to incorporate similar statistical groupings into OJS.

**Figure 4: Desired frequency of statistics capture**



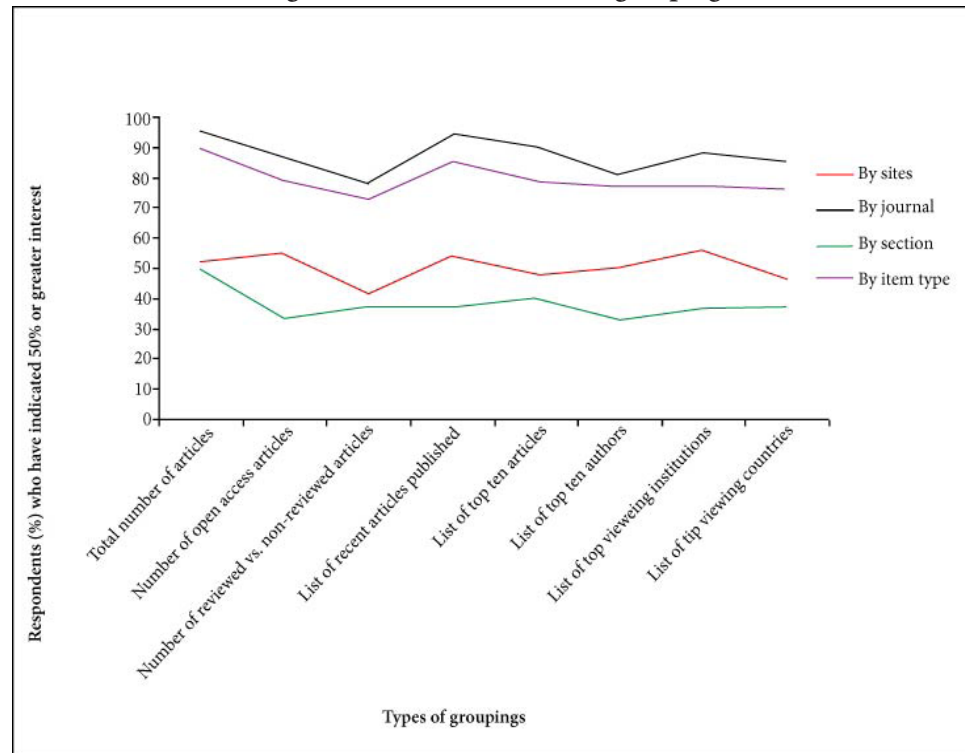
Interest in these statistical groupings was highest at the journal level, where respondents indicated that each suggested grouping was important (Figure 5). Respondents were also interested in viewing statistics grouped by item type, such as article or book review. When aggregating statistics for the Synergies portal, we hope to capture adequate statistical information to be able to display some statistical groupings at the journal level.

Respondents were also asked about current formats of statistics display and their preferences regarding display formats. The results were that 48% of respondents reported receiving statistical reports in HTML format, followed by 28% in tabulated format and 18% in PDF format. It would have been ideal for one format of statistics display to emerge as most desirable among respondents so that it would guide our choice in format for Synergies statistics display. This was not what we observed.

When asked to rank their preference for format of statistics displays, 74% of respondents ranked HTML with an importance of 50% or higher. Plain text was

favoured by 48%, while tabulated format (70%), PDF (69%), and XML (58%) were also ranked highly. Although respondents would like to continue to see the HTML format for statistics display, they are also interested in statistical information being made available in a variety of “useful” formats. From a Synergies perspective, authors viewing statistics for their article may want a simple HTML display, while journals may want to be able to download statistics in a format where they can parse and analyze them.

Figure 5: Interest in statistical groupings



#### PRELIMINARY RECOMMENDATIONS

The survey results remind us to be mindful of many different aspects of statistics collection when designing the statistical component of the Synergies platform. We are faced with the dual challenge that no single ideal statistical tool currently exists and that statistical needs vary among journals. As a result, it is unlikely that standardizing on any single statistical tool is a viable option for Synergies partners.

As the anticipated need for sharing statistics with stakeholders is higher than current practices, we need to be mindful that journal managers require the flexibility to generate customized reports that can be exported and shared with multiple stakeholders and in a variety of formats with variable permissions. This will help to ensure that barriers for the sharing of statistics are not technical in nature.

It is clear from our respondents that there is interest in a variety of statistical metrics. The aim should be to capture as many statistical types as possible with a focus on the statistics most highly demanded, such as the geographic location of visitors and the number of article downloads. The Synergies portal should aim to aggregate statistics on a monthly basis, and if funding exists to code advanced sorting capabilities for statistics, we believe users would be interested in viewing statistics grouped by article, author, and journal.



## Final recommendations

Based on the platform comparison and survey analyses, our final recommendations to the Synergies Technical Committee are as follows:

We have concluded from our survey analysis that no ideal tool exists for statistical capture. Third-party tools present unique problems: while they do generally capture interesting information, it is difficult to extract and/or interpret useful data from their default report output. As a result, attempting to standardize additional statistical tools would be a complicated and imperfect solution. Instead, we have decided to work with the statistical tools currently in use within OJS and recommend that Érudit follow suit. We intend to focus on enhancing the functionality and scope of these tools as well as developing a system of interoperability and exchange between platforms.

Currently, no readily available solution enabling the efficient transfer of statistical data between the Érudit and OJS platforms exists. This is problematic, as our survey analysis revealed that there is an anticipated greater need for sharing statistics with stakeholders, and this is also an anticipated need for the Synergies project. If Synergies statistical data were to be stored in multiple locations, this would prove to be a further obstacle to sharing data with stakeholders, as journal managers would have to find a way to aggregate these statistics on their own. In order to facilitate the sharing of statistics, developing a system that automates the aggregation of statistics between regional nodes and the head node is required.

To make possible the aggregating of statistical data, we recommend implementing the SUSHI protocol in OJS and Érudit. This will allow regional nodes to share relevant statistics with the head node, and vice versa if necessary. (Once implemented in OJS, the code implementing the SUSHI protocol can be easily added to other PKP platforms, such as the Harvester and the in-development Open Monographs Press.)

Collecting statistics is the first step toward meeting the identified anticipated need for journals to share statistics with additional stakeholders. The ability to generate reports that draw on the pool of aggregated statistics is also required. To address the need for flexibility of statistical reporting, we recommend that a custom Synergies report be developed. It should be compatible with the SUSHI protocol (and available in CSV as well as XML format), but also viewable on a per-journal or per-site basis. This custom report should include extended statistics, specifically site visits, geographic location, text of queries, item record (metadata/abstract) views, and per-item downloads. With this solution we will be able to provide users with a comprehensive report that includes the statistics they have identified as most important in our survey.

Time and funds permitting, public-facing statistics-grouping functionality in OJS (which already exists to a degree) and Érudit should be developed to address the anticipated need to share statistics with the general public.

We also recommend that, time and funds permitting, a report specifically tailored for culling usage data necessary for Social Sciences and Humanities Research Council (SSHRC) funding applications be outlined and developed. We would work with SSHRC to identify specific data points required for their assessments and would integrate their



report formatting and layout preferences. Ideally, the SSHRC report structure would be modular, where users can edit which data points are included and how they are displayed. In this way, the report could evolve as SSHRC evolves its requirements. This report should be available in csv and optionally XML format and in OJS should be used as an example for other journals to develop their own custom reports.

Following from our recommendation to make use of current tools, it is important to ensure that these tools are functioning at an optimal level. Open Journal Systems currently provides a COUNTER plug-in that is functional but not entirely compliant with the standard. We suggest implementing, both as a test case and as a baseline of useful reporting, a version of the COUNTER JR 1 report in OJS and Érudit that is compliant to Release 3 of the COUNTER Code of Practice (COUNTER, 2008). The COUNTER JR 1 report collects the number of item downloads in both HTML and PDF formats and reports these statistics on a monthly basis. Implementing COUNTER will enable reports to be made available in both XML and CSV formats; hence, correct COUNTER implementation will satisfy requests for accurate article view count reporting, in sufficiently flexible and useful report formats (XML and CSV), thus immediately satisfying many stakeholder needs.

Although the primary focus of our investigation concerns satisfying the Synergies mandate and providing useful reports to the Synergies nodes and stakeholders, we acknowledge that a substantial number of these recommendations will have an effect on other users of this software, namely journal managers and editors (and in the case of OCS, conference managers and directors). A secondary goal, then, is to create not only an infrastructure that supports the Synergies project, but also one that supports, in ways that are easily customizable and based on well-documented standards and protocols, journals, their staff, and their users, and eventually, conferences, monographs, and other forms of applicable content.

## **Conclusion**

Assessing the capture, transfer, and dissemination of statistics between scholarly journals has been a very informative experience. We found that journal stakeholders are attempting to capture and piece together data from multiple tools and sources, as no ideal tool or solution has been made available. It is crucial to be able to provide meaningful, standardized metrics that can be used as a basis of comparison for research and funding purposes. By developing this proposed framework, not only will the Synergies project benefit, but OJS users internationally will be better positioned to easily access, assess, and use the metrics they need. This is becoming increasingly necessary as journals transition from primarily print-based to primarily (or entirely) Web-based. Careful planning now, with an eye to extensibility, will ensure that these needs are met.

## **ACKNOWLEDGEMENT**

This work was researched and written primarily under the auspices of Synergies, a CFI-funded initiative. Canada Foundation for Innovation/Fondation canadienne pour l'innovation: <http://www.innovation.ca>.

## WEBSITES

Érudit Consortium. [Website]. URL: <http://erudit.org> .  
Public Knowledge Project (PKP). [Website]. URL: <http://pkp.sfu.ca> .  
Synergies. (2008). [Website]. *beta synergies*. URL: <http://synergiescanada.org> .

## REFERENCES

COUNTER (Counting Online Usage of Networked Electronic Resources). (2008, August). Code of practice for journals and databases, Release 3. URL: [http://www.projectcounter.org/code\\_practice.html](http://www.projectcounter.org/code_practice.html) .  
Mark, Linda, & Tarullo, Hope. (2003). Technology and the transformation of abstracting and indexing services. *Serials Review*, 29(3), 213-220.  
SUSHI (Standardized Usage Statistics Harvesting Initiative). (2008, August). FAQ. NISO. URL: <http://www.niso.org/workrooms/sushi> .

## Scholarly and Research Communication

VOLUME 1 / ISSUE 1 / 2010